Amendments to the Claims

Please amend claims 1, 4, 9-11, 14-16, 19, 20, 22 and cancel claims 2-3, 5-8, 12-13, and 17-18 without prejudice as follows:

Listing of the claims

1. (Currently Amended) A computer readable medium storing a computer program to perform method steps for execution by a processor~~program storage device readable by a machine, tangibly embodying a program of instructions executable by the machine to perform method steps for recognizing speech~~, the method steps comprising:

generating a synthetic waveform for each of N textual transcriptions of an original waveform, wherein N is greater than 1 and the N textual transcriptions are generated by a speech recognition system and represent N-best textual transcription hypotheses of the original waveform;

for each synthetic waveform,

time-aligning feature vectors of the synthetic waveform with feature vectors of the original waveform at a phoneme level;

computing a mean of the feature vectors which align to each phoneme for the original waveform and the synthetic waveform;

computing a distance measure between each phoneme mean of the original waveform and the synthetic waveform;

summing the distance measures to generate an overall distance measure representing a distance between the original waveform and the synthetic waveform; and

~~comparing each synthetic waveform with the original waveform decoded by the speech recognition system to determine the synthetic waveform that is closest to the original waveform; and~~

selecting for output the textual transcription corresponding to the synthetic waveform having a smallest overall distance measure~~determined to be closest to the original waveform~~.

2-3. (Cancelled)

4. (Currently Amended) The computer readable medium~~program storage device~~ of claim 1~~2~~, wherein the alignment is performed using a Viterbi alignment process.

5-8. (Cancelled)

9. (Currently Amended) A method for recognizing speech, the method comprising the steps of:

generating a synthetic waveform for each of N textual transcriptions of an original waveform, wherein N is greater than 1 and the N textual transcriptions are generated by a speech recognition system and represent N-best textual transcription hypotheses of the original waveform;

for each synthetic waveform,

computing a distance measure between the synthetic waveform and the original waveform;

summing the distance measures to generate an overall distance measure representing a distance between the original waveform and the synthetic waveform;

generating a score from the overall distance measure, an acoustic model score for the synthetic wave, and a language model score of the synthetic waveform;

~~comparing each synthetic waveform with the original waveform decoded by the speech recognition system to determine the synthetic waveform that is closest to the original waveform; and~~

selecting for output one of the textual transcriptions corresponding to the synthetic waveform ~~determined to be~~having the score that indicates the synthetic wave is closest to the original waveform.

10. (Currently Amended) The method of claim 9, ~~wherein the comparing step includes the steps of~~further comprising:

aligning frames of the original waveform and frames of each synthetic waveform to a corresponding one of the N textual transcriptions; and

calculating the distance measure between the original waveform and each of the synthetic waveforms based on the corresponding alignments.

11. (Currently Amended) The method of claim 10, ~~wherein the comparing step further includes the steps of~~further comprising:

retrieving feature vectors corresponding to the original waveform; and

generating feature vectors for each synthetic waveform such that the feature vectors for the synthetic waveforms are-similar in structure to the feature vectors of the original waveform[[;]],

wherein the alignment is performed by time-aligning the feature vectors of the original waveform and the feature vectors of each synthetic waveform with the corresponding one of the N textual transcriptions.

12-13. (Cancelled)

14. (Currently Amended) The method of claim 12~~13~~, ~~wherein the step of determining the individual distance (step d) includes the steps of~~further comprising:

computing a mean feature vector of all feature vectors comprising each aligned frame for both the original and Nth synthetic waveform, wherein the ~~individual~~ distance measure for each aligned frame is calculated by determining a distance between each means of the corresponding aligned frames.

15. (Currently Amended) An automatic speech recognition system, comprising:

a decoder for decoding an original waveform of acoustic utterances to produce N textual transcriptions, the N textual transcriptions representing N-best textual transcription hypotheses of the decoded original waveform;

~~a waveform generator~~ a text to speech system~~for~~ generating a synthetic waveform for each of the N textual transcriptions; ~~and~~

a means to perform a speaker normalization on the original waveform to match vocal-tract characteristics of a speaker from whose data the TTS is derived; and

a comparator for comparing overall distance measures between each synthetic waveform with the normalized original waveform ~~to rescore the N-best hypotheses~~ to determine a corresponding one of the N-best textual transcriptions to output,

wherein the overall distance measures are computed by:

computing a distance measure between the synthetic waveform and the normalized original waveform; and

summing the distance measures to generate an overall distance measure representing a distance between the normalized original waveform and the synthetic waveform, and

wherein N is greater than 1.

16. (Currently Amended) The system of claim 15, further comprising a feature analysis processor adapted to generate a set of feature vectors for the normalized original waveform and generate a set of feature vectors for each of the N synthetic waveforms ~~using a similar feature analysis process~~.

17-18. (Cancelled)

19. (Currently Amended) The system of claim 18, wherein the determination of the corresponding one of the N-best textual transcriptions to output is further based on~~means for determining the closest synthetic waveform utilizes one of a distance score,~~ a language model score, and an acoustic model score~~, and a combination thereof, for determining the closest distance~~.

20. (Currently Amended) The system of claim 18, ~~wherein the means for determining the closest synthetic waveform~~further comprises:

means for aligning frames of the normalized original waveform and frames of each synthetic waveform to a corresponding one of the N textual transcriptions; and

means for calculating the distance measure between the normalized original waveform and each of the synthetic waveforms based on the corresponding alignments.

21. (Original) The system of claim 20, wherein the frames are aligned on a phoneme level.

22. (Currently Amended) The system of claim 20, wherein the means for calculating the distance measures comprises:

means for calculating an individual distance between each aligned frame of the original normalized waveform and each of the N synthetic waveforms; and

means for summing the individual distances of the aligned frames of the original normalized waveform and each synthetic waveform to compute the overall distance measures between the original normalized waveform and each synthetic waveforms.